

日 本 国 特 許 庁
PATENT OFFICE
JAPANESE GOVERNMENT

JP921 U.S. PTO
09/705069
11/02/00

別紙添付の書類に記載されている事項は下記の出願書類に記載されて^ている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application:

1999年11月11日

出 願 番 号
Application Number:

平成11年特許願第321363号

出 願 人
Applicant (s):

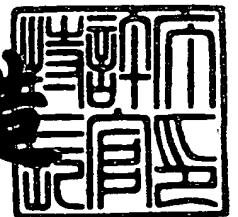
ソニー株式会社

CERTIFIED COPY OF
PRIORITY DOCUMENT

2000年 9月18日

特許庁長官
Commissioner,
Patent Office

及 川 耕 造





【書類名】 特許願
 【整理番号】 9900880606
 【提出日】 平成11年11月11日
 【あて先】 特許庁長官殿
 【国際特許分類】 G06F 7/24
 【発明者】

【住所又は居所】 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社
 内

【氏名】 安部 素嗣

【発明者】

【住所又は居所】 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社
 内

【氏名】 西口 正之

【特許出願人】

【識別番号】 000002185

【氏名又は名称】 ソニー株式会社

【代表者】 出井 伸之

【代理人】

【識別番号】 100067736

【弁理士】

【氏名又は名称】 小池 晃

【選任した代理人】

【識別番号】 100086335

【弁理士】

【氏名又は名称】 田村 榮一

【選任した代理人】

【識別番号】 100096677

【弁理士】

【氏名又は名称】 伊賀 誠司

【手数料の表示】

【予納台帳番号】 019530

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9707387

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 信号分類方法及び装置、記述子生成方法及び装置、信号検索方法及び装置

【特許請求の範囲】

【請求項 1】 入力信号を所定時間長のブロック単位に区切り、
当該ブロック内の信号より信号属性に関する特徴量を抽出し、
当該特徴量に基づいて上記ブロック単位に信号のカテゴリ分類を行う
ことを特徴とする信号分類方法。

【請求項 2】 上記ブロック内における信号の信号源の種類に基づくカテゴリ分類を行うことを特徴とする請求項 1 記載の信号分類方法。

【請求項 3】 上記ブロック内における信号が持つ信号源の種類に依存しない構造に基づくカテゴリ分類を行うことを特徴とする請求項 1 記載の信号分類方法。

【請求項 4】 上記入力信号はオーディオ信号であり、
上記ブロック内のオーディオ信号の信号源の種類に基づくカテゴリとして、無音、音声、男性音声、女性音声、音楽、ボーカル音楽、楽器音楽、雑音、打撃音、環境音、雑踏音、拍手音、歓声の、一つ又は複数を用い、音源的なカテゴリ分類を行うことを特徴とする請求項 2 記載の信号分類方法。

【請求項 5】 上記入力信号はオーディオ信号であり、
上記ブロック内のオーディオ信号の信号源の種類に依存しない構造にもとづくカテゴリとして、上記ブロック中に有意な音が存在しない無音構造、上記ブロック中に一つの音源に関する音のみが存在する単音源構造、上記ブロック中に二つの音源に関する音が存在する二音構造、上記ブロック中に無音を含む音源が一回のみ代わる音源変化構造、上記ブロック中に複数の音源が同時に変化する音源同時変化構造、上記ブロック中に複数の音源のうち一部の音源が変化する音源部分変化構造、上記ブロック中の構造が上記各構造の何れにも該当しない他の構造の、各構造のうち一つ又は複数の構造を用い、構造的なカテゴリ分類を行うことを特徴とする請求項 3 記載の信号分類方法。

【請求項 6】 上記ブロック中の信号パワーの平均及び分散、上記ブロック中の信号のバンドパス信号のパワーの平均及び分散、上記ブロック中の信号のスペクトログラムの広がり、の平均及び分散、上記ブロック中の信号のピッチ周波数の平均及び分散、上記ブロック中の信号の調波構造の程度の平均及び分散、上記ブロック中の信号の線形予測分析の残差信号の平均及び分散、上記ブロック中の信号の線形予測分析の残差信号のピッチゲインの平均及び分散の、各平均及び分散のうち、一つ又は複数の平均及び分散を、上記特徴量として用いることを特徴とする請求項 1 記載の信号分類方法。

【請求項 7】 上記調波構造の程度の平均は、上記ブロック中の信号のピッチ周波数の整数倍成分のエネルギーの全周波数エネルギーに対する比率の、上記ブロック内の時間平均であり、

上記調波構造の程度の分散は、上記ブロック中の信号のピッチ周波数の整数倍成分のエネルギーの全周波数エネルギーに対する比率の、上記ブロック内の時間的な相対標準偏差であることを特徴とする請求項 6 記載の信号分類方法。

【請求項 8】 上記カテゴリの分類方法として、ベクトル量子化法を用いることを特徴とする請求項 1 記載の信号分類方法。

【請求項 9】 入力信号を所定時間長のブロック単位に区切るブロック化部と、
当該ブロック内の信号より信号属性に関する特徴量を抽出する特徴抽出部と、
当該特徴量に基づいて上記ブロック単位に信号のカテゴリ分類を行うカテゴリ分類部とを有する
ことを特徴とする信号分類装置。

【請求項 10】 上記カテゴリ分類部は、上記ブロック内における信号の信号源の種類に基づくカテゴリ分類を行うことを特徴とする請求項 9 記載の信号分類装置。

【請求項 11】 上記カテゴリ分類部は、上記ブロック内における信号が持つ信号源の種類に依存しない構造に基づくカテゴリ分類を行うことを特徴とする請求項 9 記載の信号分類装置。

【請求項 1 2】 上記入力信号はオーディオ信号であり、

上記カテゴリ分類部は、上記ブロック内のオーディオ信号の信号源の種類に基づくカテゴリとして、無音、音声、男性音声、女性音声、音楽、ボーカル音楽、楽器音楽、雑音、打撃音、環境音、雑踏音、拍手音、歓声の、一つ又は複数を用い、音源的なカテゴリ分類を行うことを特徴とする請求項 1 0 記載の信号分類装置。

【請求項 1 3】 上記入力信号はオーディオ信号であり、

上記カテゴリ分類部は、上記ブロック内のオーディオ信号の信号源の種類に依存しない構造にもとづくカテゴリとして、上記ブロック中に有意な音が存在しない無音構造、上記ブロック中に一つの音源に関する音のみが存在する単音源構造、上記ブロック中に二つの音源に関する音が存在する二音構造、上記ブロック中に無音を含む音源が一回のみ代わる音源変化構造、上記ブロック中に複数の音源が同時に変化する音源同時変化構造、上記ブロック中に複数の音源のうち一部の音源が変化する音源部分変化構造、上記ブロック中の構造が上記各構造の何れにも該当しない他の構造の、各構造のうち一つ又は複数の構造を用い、構造的なカテゴリ分類を行うことを特徴とする請求項 1 1 記載の信号分類装置。

【請求項 1 4】 上記特徴抽出部は、上記ブロック中の信号パワーの平均及び分散、上記ブロック中の信号のバンドパス信号のパワーの平均及び分散、上記ブロック中の信号のスペクトログラムの広がり平均及び分散、上記ブロック中の信号のピッチ周波数の平均及び分散、上記ブロック中の信号の調波構造の程度の平均及び分散、上記ブロック中の信号の線形予測分析の残差信号の平均及び分散、上記ブロック中の信号の線形予測分析の残差信号のピッチゲインの平均及び分散の、各平均及び分散のうち、一つ又は複数の平均及び分散を、上記特徴量として用いることを特徴とする請求項 9 記載の信号分類装置。

【請求項 1 5】 上記調波構造の程度の平均は、上記ブロック中の信号のピッチ周波数の整数倍成分のエネルギーの全周波数エネルギーに対する比率の、上記ブロック内の時間平均であり、

上記調波構造の程度の分散は、上記ブロック中の信号のピッチ周波数の整数倍成分のエネルギーの全周波数エネルギーに対する比率の、上記ブロック内の時間的な

相対標準偏差であることを特徴とする請求項 1 4 記載の信号分類装置。

【請求項 1 6】 上記カテゴリ分類部は、上記カテゴリの分類方法として、ベクトル量子化法を用いることを特徴とする請求項 9 記載の信号分類装置。

【請求項 1 7】 入力信号を所定時間長のブロック単位に区切り、
当該ブロック内の信号より信号属性に関する特徴量を抽出し、
当該特徴量に基づいて上記ブロック単位に信号のカテゴリ分類を行い、
上記カテゴリ分類に応じた記述子を生成する
ことを特徴とする記述子生成方法。

【請求項 1 8】 上記ブロック内における信号の信号源の種類に基づくカテゴリ分類を行うことを特徴とする請求項 1 7 記載の記述子生成方法。

【請求項 1 9】 上記ブロック内における信号が持つ信号源の種類に依存しない構造に基づくカテゴリ分類を行うことを特徴とする請求項 1 7 記載の記述子生成方法。

【請求項 2 0】 上記入力信号はオーディオ信号であり、
上記ブロック内のオーディオ信号の信号源の種類に基づくカテゴリとして、無音、音声、男性音声、女性音声、音楽、ボーカル音楽、楽器音楽、雑音、打撃音、環境音、雑踏音、拍手音、歓声の、一つ又は複数を用い、音源的なカテゴリ分類を行い、

上記音源的なカテゴリ分類に応じた記述子を生成することを特徴とする請求項 1 8 記載の記述子生成方法。

【請求項 2 1】 上記入力信号はオーディオ信号であり、
上記ブロック内のオーディオ信号の信号源の種類に依存しない構造にもとづくカテゴリとして、上記ブロック中に有意な音が存在しない無音構造、上記ブロック中に一つの音源に関する音のみが存在する単音源構造、上記ブロック中に二つの音源に関する音が存在する二音構造、上記ブロック中に無音を含む音源が一回のみ代わる音源変化構造、上記ブロック中に複数の音源が同時に変化する音源同時変化構造、上記ブロック中に複数の音源のうち一部の音源が変化する音源部分変化構造、上記ブロック中の構造が上記各構造の何れにも該当しない他の構造の、各構造のうち一つ又は複数の構造を用い、構造的なカテゴリ分類を行い、

上記構造的なカテゴリ分類に応じた記述子を生成することを特徴とする請求項 1 9 記載の記述子生成方法。

【請求項 2 2】 上記ブロック中の信号パワーの平均及び分散、上記ブロック中の信号のバンドパス信号のパワーの平均及び分散、上記ブロック中の信号のスペクトログラムの広がり、の平均及び分散、上記ブロック中の信号のピッチ周波数の平均及び分散、上記ブロック中の信号の調波構造の程度の平均及び分散、上記ブロック中の信号の線形予測分析の残差信号の平均及び分散、上記ブロック中の信号の線形予測分析の残差信号のピッチゲインの平均及び分散の、各平均及び分散のうち、一つ又は複数の平均及び分散を、上記特徴量として用いることを特徴とする請求項 1 7 記載の記述子生成方法。

【請求項 2 3】 上記調波構造の程度の平均は、上記ブロック中の信号のピッチ周波数の整数倍成分のエネルギーの全周波数エネルギーに対する比率の、上記ブロック内の時間平均であり、

上記調波構造の程度の分散は、上記ブロック中の信号のピッチ周波数の整数倍成分のエネルギーの全周波数エネルギーに対する比率の、上記ブロック内の時間的な相対標準偏差であることを特徴とする請求項 2 2 記載の記述子生成方法。

【請求項 2 4】 上記カテゴリの分類方法として、ベクトル量子化法を用いることを特徴とする請求項 1 7 記載の記述子生成方法。

【請求項 2 5】 入力信号を所定時間長のブロック単位に区切るブロック化部と、

当該ブロック内の信号より信号属性に関する特徴量を抽出する特徴抽出部と、

当該特徴量に基づいて上記ブロック単位に信号のカテゴリ分類を行うカテゴリ分類部と、

上記カテゴリ分類に応じた記述子を生成する記述子生成部とを有する

ことを特徴とする記述子生成装置。

【請求項 2 6】 上記カテゴリ分類部は、上記ブロック内における信号の信号源の種類に基づくカテゴリ分類を行うことを特徴とする請求項 2 5 記載の記述子生成装置。

【請求項 2 7】 上記カテゴリ分類部は、上記ブロック内における信号が持つ信号源の種類に依存しない構造に基づくカテゴリ分類を行うことを特徴とする請求項 2 5 記載の記述子生成装置。

【請求項 2 8】 上記入力信号はオーディオ信号であり、

上記カテゴリ分類部は、上記ブロック内のオーディオ信号の信号源の種類に基づくカテゴリとして、無音、音声、男性音声、女性音声、音楽、ボーカル音楽、楽器音楽、雑音、打撃音、環境音、雑踏音、拍手音、歓声の、一つ又は複数を用い、音源的なカテゴリ分類を行い、

上記記述子生成部は、上記音源的なカテゴリ分類に応じた記述子を生成することを特徴とする請求項 2 6 記載の記述子生成装置。

【請求項 2 9】 上記入力信号はオーディオ信号であり、

上記カテゴリ分類部は、上記ブロック内のオーディオ信号の信号源の種類に依存しない構造にもとづくカテゴリとして、上記ブロック中に有意な音が存在しない無音構造、上記ブロック中に一つの音源に関する音のみが存在する単音源構造、上記ブロック中に二つの音源に関する音が存在する二音構造、上記ブロック中に無音を含む音源が一回のみ代わる音源変化構造、上記ブロック中に複数の音源が同時に変化する音源同時変化構造、上記ブロック中に複数の音源のうち一部の音源が変化する音源部分変化構造、上記ブロック中の構造が上記各構造の何れにも該当しない他の構造の、各構造のうち一つ又は複数の構造を用い、構造的なカテゴリ分類を行い、

上記記述子生成部は、上記構造的なカテゴリ分類に応じた記述子を生成することを特徴とする請求項 2 7 記載の記述子生成装置。

【請求項 3 0】 上記特徴抽出部は、上記ブロック中の信号パワーの平均及び分散、上記ブロック中の信号のバンドパス信号のパワーの平均及び分散、上記ブロック中の信号のスペクトログラムの広がり、の平均及び分散、上記ブロック中の信号のピッチ周波数の平均及び分散、上記ブロック中の信号の調波構造の程度の平均及び分散、上記ブロック中の信号の線形予測分析の残差信号の平均及び分散、上記ブロック中の信号の線形予測分析の残差信号のピッチゲインの平均及び分散の、各平均及び分散のうち、一つ又は複数の平均及び分散を、上記特徴量とし

て用いることを特徴とする請求項 2 5 記載の記述子生成装置。

【請求項 3 1】 上記調波構造の程度の平均は、上記ブロック中の信号のピッチ周波数の整数倍成分のエネルギーの全周波数エネルギーに対する比率の、上記ブロック内の時間平均であり、

上記調波構造の程度の分散は、上記ブロック中の信号のピッチ周波数の整数倍成分のエネルギーの全周波数エネルギーに対する比率の、上記ブロック内の時間的な相対標準偏差であることを特徴とする請求項 3 0 記載の記述子生成装置。

【請求項 3 2】 上記カテゴリ分類部は、上記カテゴリの分類方法として、ベクトル量子化法を用いることを特徴とする請求項 2 5 記載の記述子生成装置。

【請求項 3 3】 入力信号を所定時間長のブロック単位に区切り、
当該ブロック内の信号より信号属性に関する特徴量を抽出し、
当該特徴量に基づいて上記ブロック単位に信号のカテゴリ分類を行い、
上記カテゴリ分類結果或いは当該カテゴリ分類結果に応じた記述子を用いて信号検索を行う

ことを特徴とする信号検索方法。

【請求項 3 4】 上記ブロック内における信号の信号源の種類に基づくカテゴリ分類を行うことを特徴とする請求項 3 3 記載の信号検索方法。

【請求項 3 5】 上記ブロック内における信号が持つ信号源の種類に依存しない構造に基づくカテゴリ分類を行うことを特徴とする請求項 3 3 記載の信号検索方法。

【請求項 3 6】 上記入力信号はオーディオ信号であり、

上記ブロック内のオーディオ信号の信号源の種類に基づくカテゴリとして、無音、音声、男性音声、女性音声、音楽、ボーカル音楽、楽器音楽、雑音、打撃音、環境音、雑踏音、拍手音、歓声の、一つ又は複数を用い、音源的なカテゴリ分類を行い、

上記音源的なカテゴリ分類結果或いは当該音源的なカテゴリ分類結果に応じた記述子を用いて信号検索を行うことを特徴とする請求項 3 4 記載の信号検索方法

【請求項 3 7】 上記入力信号はオーディオ信号であり、

上記ブロック内のオーディオ信号の信号源の種類に依存しない構造にもとづくカテゴリとして、上記ブロック中に有意な音が存在しない無音構造、上記ブロック中に一つの音源に関する音のみが存在する単音源構造、上記ブロック中に二つの音源に関する音が存在する二音構造、上記ブロック中に無音を含む音源が一回のみ代わる音源変化構造、上記ブロック中に複数の音源が同時に変化する音源同時変化構造、上記ブロック中に複数の音源のうち一部の音源が変化する音源部分変化構造、上記ブロック中の構造が上記各構造の何れにも該当しない他の構造の、各構造のうち一つ又は複数の構造を用い、構造的なカテゴリ分類を行い、

上記構造的なカテゴリ分類結果或いは当該構造的なカテゴリ分類結果に応じた記述子を用いて信号検索を行うことを特徴とする請求項 3 5 記載の信号検索方法。

【請求項 3 8】 上記ブロック中の信号パワーの平均及び分散、上記ブロック中の信号のバンドパス信号のパワーの平均及び分散、上記ブロック中の信号のスペクトログラムの広がり平均及び分散、上記ブロック中の信号のピッチ周波数の平均及び分散、上記ブロック中の信号の調波構造の程度の平均及び分散、上記ブロック中の信号の線形予測分析の残差信号の平均及び分散、上記ブロック中の信号の線形予測分析の残差信号のピッチゲインの平均及び分散の、各平均及び分散のうち、一つ又は複数の平均及び分散を、上記特徴量として用いることを特徴とする請求項 3 3 記載の信号検索方法。

【請求項 3 9】 上記調波構造の程度の平均は、上記ブロック中の信号のピッチ周波数の整数倍成分のエネルギーの全周波数エネルギーに対する比率の、上記ブロック内の時間平均であり、

上記調波構造の程度の分散は、上記ブロック中の信号のピッチ周波数の整数倍成分のエネルギーの全周波数エネルギーに対する比率の、上記ブロック内の時間的な相対標準偏差であることを特徴とする請求項 3 8 記載の信号検索方法。

【請求項 4 0】 上記カテゴリの分類方法として、ベクトル量子化法を用いることを特徴とする請求項 3 3 記載の信号検索方法。

【請求項 4 1】 上記カテゴリ分類結果或いは当該カテゴリ分類結果に応じた記述子を用いて、信号の変化点を検索することを特徴とする請求項 3 3 記載の信号検索方法。

【請求項 4 2】 入力信号を所定時間長のブロック単位に区切るブロック化部と、

当該ブロック内の信号より信号属性に関する特徴量を抽出する特徴抽出部と、
当該特徴量に基づいて上記ブロック単位に信号のカテゴリ分類を行うカテゴリ分類部と、

上記カテゴリ分類結果或いは当該カテゴリ分類に応じた記述子を用いて、信号検索を行う信号検索部とを有する

ことを特徴とする信号検索装置。

【請求項 4 3】 上記カテゴリ分類部は、上記ブロック内における信号の信号源の種類に基づくカテゴリ分類を行うことを特徴とする請求項 4 2 記載の信号検索装置。

【請求項 4 4】 上記カテゴリ分類部は、上記ブロック内における信号が持つ信号源の種類に依存しない構造に基づくカテゴリ分類を行うことを特徴とする請求項 4 2 記載の信号検索装置。

【請求項 4 5】 上記入力信号はオーディオ信号であり、

上記カテゴリ分類部は、上記ブロック内のオーディオ信号の信号源の種類に基づくカテゴリとして、無音、音声、男性音声、女性音声、音楽、ボーカル音楽、楽器音楽、雑音、打撃音、環境音、雑踏音、拍手音、歓声の、一つ又は複数を用い、音源的なカテゴリ分類を行い、

上記信号検索部は、上記音源的なカテゴリ分類結果或いは当該音源的なカテゴリ分類結果に応じた記述子を用いて信号検索を行うことを特徴とする請求項 4 3 記載の信号検索装置。

【請求項 4 6】 上記入力信号はオーディオ信号であり、

上記カテゴリ分類部は、上記ブロック内のオーディオ信号の信号源の種類に依存しない構造にもとづくカテゴリとして、上記ブロック中に有意な音が存在しない無音構造、上記ブロック中に一つの音源に関する音のみが存在する単音源構造

、上記ブロック中に二つの音源に関する音が存在する二音構造、上記ブロック中に無音を含む音源が一回のみ代わる音源変化構造、上記ブロック中に複数の音源が同時に変化する音源同時変化構造、上記ブロック中に複数の音源のうち一部の音源が変化する音源部分変化構造、上記ブロック中の構造が上記各構造の何れにも該当しない他の構造の、各構造のうち一つ又は複数の構造を用い、構造的なカテゴリ分類を行い、

上記信号検索部は、上記構造的なカテゴリ分類結果或いは当該構造的なカテゴリ分類結果に応じた記述子を用いて信号検索を行うことを特徴とする請求項 4 4 記載の信号検索装置。

【請求項 4 7】 上記特徴抽出部は、上記ブロック中の信号パワーの平均及び分散、上記ブロック中の信号のバンドパス信号のパワーの平均及び分散、上記ブロック中の信号のスペクトログラムの広がり、の平均及び分散、上記ブロック中の信号のピッチ周波数の平均及び分散、上記ブロック中の信号の調波構造の程度の平均及び分散、上記ブロック中の信号の線形予測分析の残差信号の平均及び分散、上記ブロック中の信号の線形予測分析の残差信号のピッチゲインの平均及び分散の、各平均及び分散のうち、一つ又は複数の平均及び分散を、上記特徴量として用いることを特徴とする請求項 4 2 記載の信号検索装置。

【請求項 4 8】 上記調波構造の程度の平均は、上記ブロック中の信号のピッチ周波数の整数倍成分のエネルギーの全周波数エネルギーに対する比率の、上記ブロック内の時間平均であり、

上記調波構造の程度の分散は、上記ブロック中の信号のピッチ周波数の整数倍成分のエネルギーの全周波数エネルギーに対する比率の、上記ブロック内の時間的な相対標準偏差であることを特徴とする請求項 4 7 記載の信号検索装置。

【請求項 4 9】 上記カテゴリ分類部は、上記カテゴリの分類方法として、ベクトル量子化法を用いることを特徴とする請求項 4 2 記載の信号検索装置。

【請求項 5 0】 上記信号検索部は、上記カテゴリ分類結果或いは当該カテゴリ分類結果に応じた記述子を用いて、信号の変化点を検索することを特徴とする請求項 4 2 記載の信号検索装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、映像信号やオーディオ信号などのマルチメディア情報を効率良く分類するための信号分類方法及び装置、その分類に応じた記述子（タグ）を生成する記述子作成方法及び装置、さらにその分類結果或いは記述子に基づいて入力信号の検索を行う信号検索方法及び装置に関する。

【0002】

【従来の技術】

従来より、映像信号やオーディオ信号などのマルチメディア情報を効率良く取り扱うためには、それら映像信号やオーディオ信号をその信号の内容に基づいて分類し、当該信号の内容に基づいた属性情報（タグ）をそれら信号に付与することが必要であるとされている。

【0003】

【発明が解決しようとする課題】

以下、上記マルチメディア情報のうち、オーディオ信号を例に挙げ、当該オーディオ信号の内容に基づく従来の分類手法について簡単に説明する。

【0004】

一般に、オーディオ信号は、音声等が存在する音声区間と音声等が存在しない無音区間とからなっている。このため、従来より、オーディオ信号の時々刻々変化する属性を分類する手段の一つとして、当該オーディオ信号の無音区間検出を行うことが多い。上記検出された無音区間の信号には、当該区間が無音区間であることを示すタグが付与され、後段の音声信号処理では、当該タグが付与された無音区間についての信号処理を停止する等の制御が行われる。

【0005】

また、オーディオ信号の分類手法として、特開平 1 0 - 2 0 7 4 9 1 号公報には、背景音と音声进行分类する手法が開示されている。当該公報記載の分類手法は、背景音のパワー及びスペクトルを推定し、当該背景音のパワー及びスペクトルと入力信号のパワー及びスペクトルとを比較することで、入力信号から背景音区

間と音声区間を分類するものである。

【0006】

しかしながら、当該公報記載の分類手法は、入力信号が音声信号であり且つ背景音が比較的定常な場合には有効に働くとされているものの、入力信号が例えば音楽、音響信号等の一般的なオーディオ信号を含む場合に、それらの分類を行い得るものではない。

【0007】

また、入力信号が音声、音楽、音響信号等のオーディオ信号を含む場合にそれらの分類を可能とする技術として、特開平 1 0 - 1 8 7 1 2 8 号公報には、映像情報に付随する音情報から映像の種類を判別する映像分類方法が開示されている。当該公報の分類手法では、先ず、入力信号から所定のスペクトル構造を有する信号を音楽信号として分類し、次に、当該入力信号から上記音楽信号の成分を除去し、その残差信号から更に所定のスペクトル構造を有する信号を音声信号として分類し、次に、上記残差信号から上記音声信号の成分を除去し、その残差信号から更に所定のスペクトル構造を有する信号を音響信号として分類するようにしている。

【0008】

しかしながら、当該公報記載の分類手法は、線スペクトル構造が一定して続く区間のみを音楽信号として取り扱うため、例えば打楽器や歌声などの信号を含む音楽信号に対しては安定して適用することが難しく、また、音声区間は入力信号の原スペクトルと安定した線スペクトル成分（音楽成分）との残差から求めるため、音声及び音響信号の判別精度が十分には高くはなり得ないと推測される。

【0009】

そこで、本発明はこのような状況に鑑みてなされたものであり、入力信号を正確且つ効率良く取り扱うために、入力信号を迅速且つ正確に分類し、またその分類に応じたタグ（記述子）を生成可能とし、さらにその分類結果或いは記述子に基づいて入力信号の検索をも可能とする、信号分類方法及び装置、記述子生成方法及び装置、信号検索方法及び装置を提供することを目的とする。

【0 0 1 0】

【課題を解決するための手段】

本発明の信号分類方法は、入力信号を所定時間長のブロック単位に区切り、当該ブロック内の信号より信号属性に関する特徴量を抽出し、当該特徴量に基づいて上記ブロック単位に信号のカテゴリ分類を行うことにより、上述した課題を解決する。

【0 0 1 1】

本発明の信号分類装置は、入力信号を所定時間長のブロック単位に区切るブロック化部と、当該ブロック内の信号より信号属性に関する特徴量を抽出する特徴抽出部と、当該特徴量に基づいて上記ブロック単位に信号のカテゴリ分類を行うカテゴリ分類部とを有することにより、上述した課題を解決する。

【0 0 1 2】

本発明の記述子生成方法は、入力信号を所定時間長のブロック単位に区切り、当該ブロック内の信号より信号属性に関する特徴量を抽出し、当該特徴量に基づいて上記ブロック単位に信号のカテゴリ分類を行い、上記カテゴリ分類に応じた記述子を生成することにより、上述した課題を解決する。

【0 0 1 3】

本発明の記述子生成装置は、入力信号を所定時間長のブロック単位に区切るブロック化部と、当該ブロック内の信号より信号属性に関する特徴量を抽出する特徴抽出部と、当該特徴量に基づいて上記ブロック単位に信号のカテゴリ分類を行うカテゴリ分類部と、上記カテゴリ分類に応じた記述子を生成する記述子生成部とを有することにより、上述した課題を解決する。

【0 0 1 4】

本発明の信号検索方法は、入力信号を所定時間長のブロック単位に区切り、当該ブロック内の信号より信号属性に関する特徴量を抽出し、当該特徴量に基づいて上記ブロック単位に信号のカテゴリ分類を行い、上記カテゴリ分類結果或いは当該カテゴリ分類結果に応じた記述子を用いて信号検索を行うことにより、上述した課題を解決する。

【0015】

本発明の信号検索装置は、入力信号を所定時間長のブロック単位に区切るブロック化部と、当該ブロック内の信号より信号属性に関する特徴量を抽出する特徴抽出部と、当該特徴量に基づいて上記ブロック単位に信号のカテゴリ分類を行うカテゴリ分類部と、上記カテゴリ分類結果或いは当該カテゴリ分類に応じた記述子を用いて信号検索を行う信号検索部とを有することにより、上述した課題を解決する。

【0016】

すなわち、本発明は、長時間に渡り連続して入力される信号を所定時間長のブロック単位毎にブロック化し、該ブロック内の信号より特徴量を抽出し、該特徴量に応じて自動的に該ブロック中の信号の分類を行うものである。ここで、本発明は、入力信号がオーディオ信号である場合の分類方式として、音声、音楽、環境音等の音源に関連する音源的な分類と、無音、単一音源音、二音源音、音源変化音等の個別の音源に依存せずにブロック中の音の存在や重なり方、続き方に関連する構造的な分類とを行うようにしている。このように、本発明においては、上記構造的な分類と音源的な分類を併用することで、連続して生じる様々な音響シーンについて、頑強かつ効率的な分類を行うことが可能となっている。なお、上記ブロック化の所定時間長は、上記ブロック内の信号属性が明瞭に識別でき、且つ、上記ブロック内の信号構造がシンプルに分類可能となる時間長であり、例えば1秒とすることが好ましい。もちろん、本発明において上記ブロック化の時間長は当該1秒に限定されず、他の時間長でも良い。また、各ブロックの時間長は上記1秒のような一定の時間長に限らず、ブロック毎に可変の時間長とすること、例えば複数の時間長を切り替えて用いたり、適応的に時間長を可変することなども可能であり、それらも本発明に含まれる。

【0017】

【発明の実施の形態】

本発明の好ましい実施の形態について、図面を参照しながら説明する。

【0018】

図1には、本発明の第1の実施の形態として、入力信号（例えばオーディオ信

号) の分類を行う信号処理装置の概略構成を示す。

【 0 0 1 9 】

この図 1 において、入力端子 1 にはオーディオ信号が入力される。当該オーディオ信号は、バッファメモリ 2 に一旦蓄えられた後に読み出され、時区間分割部 3 に送られる。

【 0 0 2 0 】

時区間分割部 3 では、上記バッファメモリ 2 から供給されたオーディオ信号を所定時間長のブロック単位毎にブロック化（時区間分割）し、そのブロック化されたオーディオ信号を特徴抽出部 4 に送る。なお、当該時区間分割部 3 でのブロック化の詳細については後述する。

【 0 0 2 1 】

特徴抽出部 4 は、上記ブロック化されたオーディオ信号から、後述するような複数の特徴量を抽出し、それら複数の特徴量をベクトル量子化部 5 に送る。なお、当該特徴抽出部 4 での特徴抽出処理の詳細については後述する。

【 0 0 2 2 】

ベクトル量子化部 5 は、詳細については後述するが、いわゆるベクトル量子化法を使用し、上記特徴抽出ブロック 4 から供給された複数の特徴量からなるベクトル（以下、特徴ベクトルと呼ぶ）と、予め学習により生成された複数のセントロイド（パターン空間内の重心）の集合である VQ 符号帳（ベクトル量子化のコードブック）8 との比較（パターンマッチング）を行い、上記特徴ベクトルに最もマハラノビス距離の近いセントロイドを探索し、当該最も近いセントロイドが代表している代表符号を出力する。すなわち本実施の形態の場合、上記ベクトル量子化部 5 から出力される代表符号は、オーディオ信号を後述する音源的な分類カテゴリと構造的な分類カテゴリに分けた場合における各分類カテゴリに対応した分類ラベルとなっている。言い換えれば、当該ベクトル量子化部 5 からは、特徴ベクトルに基づいてオーディオ信号を分類した分類結果が出力されることになる。また、本実施の形態の場合、ベクトル量子化部 5 からは、上記分類ラベルと共に、上記探索されたセントロイドに対する最小距離の逆数をそのカテゴリに対する分類の信頼度を表す指標として出力するようにしている。上述したように、

本実施の形態のベクトル量子化部 5 から出力された上記構造的な分類による分類ラベルとその信頼度、及び、上記音源的な分類による分類ラベルとその信頼度は、端子 6 から分類結果を表す信号記述子として出力される。

【0023】

以下、図 1 の時区間分割部 3 における入力オーディオ信号のブロック化、特徴抽出部 4 におけるブロック毎のオーディオ信号の特徴抽出、ベクトル量子化法を用いたベクトル量子化部 5 におけるブロック毎のオーディオ信号の分類について個々に詳細に説明する。

【0024】

先ず、図 1 の時区間分割部 3 におけるオーディオ信号のブロック化（時区間分割：Time Block Division）から説明する。

【0025】

時区間分割部 3 では、以後に行われる分類操作を容易にするために、一般的に長時間に渡り様々な音の時系列として与えられるオーディオ信号を適切な時間区間に区切るようにしている。

【0026】

ここで、例えば、数十秒にも渡るオーディオ信号を何らかのカテゴリに対応させることは、その信号が内包する音の種類やパターンを非常に複雑化させてしまい、その一方で、音の分類を行う場合、信号の時間変化パターンも重要な要因であるため、例えば数十ミリ秒の信号素片のみから、音声／音楽／雑音などのカテゴリを精度良く判別することは難しい。

【0027】

そこで、本実施の形態では、信号を分割する際の時間長の要件として、「区切られた信号素片の属性が精度良く識別できること」、「区切られた信号素片の構造が単純に分類可能であること」を考慮し、上記時区間分割部 3 における分割時間を 1 秒としている。

【0028】

また、本実施の形態では、以後に行われる分類操作の精度を高めるために、図 2 に示すように、分割区間長の半分の時間で各ブロックがオーバーラップするよ

うな時区間分割を行うようにしている。すなわち、本実施の形態の時区間分割部 3 は、図 2 において B 0, B 1, B 2, B 3, . . . にて示す各ブロックの時間長をそれぞれ 1 秒にすると共に、B 0 ブロックの後半に B 1 ブロックの前半がオーバーラップし、B 1 ブロックの後半に B 2 ブロックの前半がオーバーラップし、B 2 ブロックの後半に B 3 ブロックの前半がオーバーラップし、B 3 ブロックの後半に B 4 ブロックの前半がオーバーラップするように、オーディオ信号をそれぞれブロック化するようにしている。

【0029】

次に、図 1 の特徴抽出部 4 における各ブロック毎の信号の特徴抽出 (Feature Extraction) について説明する。

【0030】

特徴抽出部 4 では、上記時区間分割部 3 にて分割された各ブロック毎に、そのブロック内の信号から後述する分類に適した特徴量を抽出するようにしている。

【0031】

以下、当該特徴抽出部 4 が各ブロック毎に抽出する特徴量について詳細に説明する。なお、以下の説明のために用いる各記号において、 t は時間を表す変数、 T はブロックの長さ (= 1 秒)、 i はブロックの番号、 $s_i(t)$ は第 i ブロックの信号 (但し $0 \leq t \leq T$)、 ω は周波数を表す変数、 Ω は周波数上限値 (本発明の処理を離散的に実現する場合、サンプリング周波数の半分がこれにあたる)、 $S_i(t, \omega)$ は第 i ブロックの信号のスペクトログラム (但し $0 \leq t \leq T$ 、 $0 \leq \omega \leq \Omega$)、 $E[]$ はブロックに渡る時間平均値、 $V[]$ はブロックに渡る時間的な相対標準偏差 (分散の自乗根を平均値で規格化した値) をそれぞれ意味する。

【0032】

上記特徴抽出部 4 では、各ブロック毎の信号の特徴量として、以下に述べるように、ブロック中の信号パワーの平均 P_m 及び標準偏差 P_{sd} 、ブロック中の信号のスペクトログラムの広がり (平均 W_m 及び標準偏差 W_{sd})、ブロック中の信号の低周波成分のパワーの平均 L_m 及び標準偏差 L_{sd} 、ブロック中の信号の中間周波成分のパワーの平均 M_m 及び標準偏差 M_{sd} 、ブロック中の信号の高周波成分のパワーの平均 H_m 及び標準偏差 H_{sd} 、ブロック中の信号のピッチ周波数の平均 F_m 及

び標準偏差 F_{sd} 、ブロック中の信号の調波構造の程度（調波構造度）の平均 A_m 及び標準偏差 A_{sd} 、ブロック中の信号の線形予測分析（LPC）の残差エネルギーの平均 R_m 及び標準偏差 R_{sd} 、ブロック中の信号のLPCの残差信号のピッチゲインの平均 G_m 及び標準偏差 G_{sd} の合計 18 種類の特徴量を求める。

【0033】

ここで、上記ブロック中の信号パワーの平均 P_m 及び標準偏差 P_{sd} は、下記式（1）、式（2）にて表される。

【0034】

$$P_m = E[s^2(t)] \quad (1)$$

$$P_{sd} = V[s^2(t)] \quad (2)$$

上記ブロック中の信号のスペクトログラムの広がり（平均 W_m 及び標準偏差 W_{sd} ）は、下記式（3）、式（4）にて表される。なお、本実施の形態では、スペクトルとして 16 kHz のサンプリング周波数による 512 点のサンプル（31.25 ms 毎のサンプル）を用いている。

【0035】

$$W_m = E[w(t)] \quad (3)$$

$$W_{sd} = V[w(t)] \quad (4)$$

但し、 $w(t)$ は下記式（5）にて表され、スペクトログラムの大きさが一定閾値を越える周波数幅を表す。特に、音楽の場合、上記 $w(t)$ で示される周波数幅は広く且つ一定になる傾向が強く、一方、音声の場合、上記 $w(t)$ で示される周波数幅は一定せず、バラバラになる傾向が強い。したがって、当該 $w(t)$ で示される周波数幅は、音楽と音声とそれ以外の音の一つの特徴量を表すものとして使用できる。

【 0 0 3 6 】

【数 1】

$$w(t) = \frac{1}{\Omega} \int_{\Gamma} d\omega, \Gamma = \left\{ \omega \mid \text{Si}(t, \omega) > \text{Threshold} \right\} \quad \cdot \cdot \cdot (5)$$

【 0 0 3 7 】

上記ブロック中の信号の低周波成分のパワーの平均 L_m 及び標準偏差 L_{sd} は、下記式 (6)、式 (7) にて表される。なお、本実施の形態では、上記低周波成分として 0 ～ 70 Hz を使用する。

【 0 0 3 8 】

$$W_m = E [1(t)] \quad (6)$$

$$W_{sd} = V [1(t)] \quad (7)$$

但し、 $1(t)$ は下記式 (8) にて表され、各時刻での信号の低周波成分の規格化されたパワーを表す。特に、音声は 70 Hz 以下の周波数成分を殆ど持たず、ドラム等の打楽器の音は 70 Hz 以下の周波数成分を多く含む。したがって、この低周波成分は、音声と打楽器音とそれ以外の音の一つの特徴量を表すものとして使用できる。

【 0 0 3 9 】

【数 2】

$$l(t) = \frac{\int_{\omega_0}^{\omega_1} S_i(t, \omega) d\omega}{\int_0^{\omega} S_i(t, \omega) d\omega}, \left(\omega_0 = 0 \text{ Hz}, \omega_1 = 70 \text{ Hz} \right) \quad \cdot \cdot \cdot (8)$$

【 0 0 4 0 】

上記ブロック中の信号の中間周波成分のパワーの平均 M_m 及び標準偏差 M_{sd} は下記式(9)、式(10)にて表される。なお、本実施の形態では、上記低周波成分として70Hz～4kHzを使用する。

【 0 0 4 1 】

$$M_m = E[m(t)] \quad (9)$$

$$M_{sd} = V[m(t)] \quad (10)$$

但し、 $m(t)$ は下記式(11)にて表され、音声帯域を略々カバーするバンドパス成分の規格化されたパワーを表す。特に、音声の周波数は、殆どが70Hz～4kHzの周波数帯域内に含まれ、したがって、当該中間周波成分は、音声とそれ以外の一つの特徴量を表すものとして使用できる。

【 0 0 4 2 】

【数3】

$$m(t) = \frac{\int_{\omega_1}^{\omega_2} S_i(t, \omega) d\omega}{\int_0^n S_i(t, \omega) d\omega}, \quad (\omega_1 = 70\text{Hz}, \omega_2 = 4\text{kHz}) \quad \cdot \cdot \cdot (11)$$

【 0 0 4 3 】

上記ブロック中の信号の高周波成分のパワーの平均 H_m 及び標準偏差 H_{sd} は下記式(12)、式(13)にて表される。なお、本実施の形態では、上記低周波成分として4kHz～8kHzを使用する。

【 0 0 4 4 】

$$H_m = E[h(t)] \quad (12)$$

$$H_{sd} = V[h(t)] \quad (13)$$

但し、 $h(t)$ は下記式(14)にて表され、信号の高周波成分の規格化されたパワーを表す。特に、音声は4kHz以上の周波数成分を殆ど持たず、例えばシンバル音等のパーカッション系の楽器音は4kHz～8kHzの周波数成分を多

く含む。したがって、この高周波成分は、音声とパーカッション系の楽器音とそれ以外の音の一つの特徴量を表すものとして使用できる。

【0045】

【数4】

$$h(t) = \frac{\int_{\omega_2}^{\omega_3} S_i(t, \omega) d\omega}{\int_0^u S_i(t, \omega) d\omega}, \left(\omega_2 = 4kHz, \omega_3 = 8kHz, \right) \quad \cdot \cdot \cdot (14)$$

【0046】

上記ブロック中の信号のピッチ周波数の平均 F_m 及び標準偏差 F_{sd} は下記式 (15)、式 (16) にて表される。

【0047】

$$F_m = E[f(t)] \quad (15)$$

$$F_{sd} = V[f(t)] \quad (16)$$

但し、 $f(t)$ は時刻 t における信号のピッチ周波数であり、例えばパーソン (Parson) の手法 (T. Parson: Separation of Speech from Interfering Speech by ms of Harmonic Selection, J. Acoust. Soc. Am., 60, 4, 911/918 (1976)) に従って求める。特に、ピッチ周波数は、後述の調波構造度による特徴抽出に使用すると共に、音楽と音声、さらには男性音性、女性音声などでそれぞれピッチ周波数が一般に異なるため、それらの音の一つの特徴量を表すものとして使用できる。

【0048】

上記ブロック中の信号の調波構造の程度 (本実施の形態では調波構造度 $a(t)$ と呼んでいる。) の平均 A_m 及び標準偏差 A_{sd} は下記式 (17)、式 (18) にて表される。

【0049】

$$A_m = E[a(t)] \quad (17)$$

$$A_{sd} = V[a(t)] \quad (18)$$

但し、 $a(t)$ は下記式(19)にて表され、これはピッチ周波数の整数倍音成分のエネルギーの全周波数エネルギーに対する比率である。また、 Δ は例えば ± 15 Hz等の微小周波数を表す。特に、調波構造度はノイズ状の音の時に著しく低くなる。したがって、当該調波構造度は、ノイズ状の音とそれ以外の音の一つの特徴量を表すものとして使用できる。

【0050】

【数5】

$$a(t) = \frac{\int_r S_i(t, \omega) d\omega}{\int S_i(t, \omega) d\omega}, \Gamma = \left\{ \omega \mid nf(t) - \Delta \leq \omega \leq nf(t) + \Delta, n = 1, 2, \dots \right\}$$

・ ・ ・ (19)

【0051】

上記ブロック中の信号の線形予測分析(LPC)の残差エネルギーの平均 R_m 及び標準偏差 R_{sd} は下記式(20)、式(21)にて表される。

【0052】

$$R_m = E[r^2(t)] / E[s^2(t)] \quad (20)$$

$$R_{sd} = V[r^2(t)] / V[s^2(t)] \quad (21)$$

但し、 $r(t)$ はLPC分析(例えば30msフレーム、12次で行う)の残差信号である。これは、当該ブロック内のスペクトル構造の複雑さ(ノイズや子音など)に関する評価量であり、原信号パワーの平均及び標準偏差に対する比率として求める。したがって、当該LPC残差エネルギーは、ノイズや子音とそれ以外の音の一つの特徴量を表すものとして使用できる。

【0053】

上記ブロック中の信号のLPCの残差信号のピッチゲインの平均 G_m 及び標準偏差 G_{sd} は下記式(22)、式(23)にて表される。

【0054】

$$G_m = E[g(t)] \quad (22)$$

$$G_{sd} = V[g(t)] \quad (23)$$

但し、 $g(t)$ は $r(t)$ の時刻 t 近傍における短時間自己相関関数の極大値であり、当該ブロック内のLPC分析（例えば30msフレーム、12次で行う）の残差信号の周期性の程度に関する評価量である。特に、LPC残差信号のピッチゲインは、ホワイトノイズや子音などの音では著しく低い値となり、一方、音声や音楽などの音では高い値となる。したがって、当該LPC残差信号のピッチゲインは、ノイズや子音と音声や音楽とそれ以外の音の一つの特徴量を表すものとして使用できる。

【0055】

本実施の形態では、以上説明した18種類の特徴量を、次式(24)のようにベクトル化し、次式(24)で表される当該ブロック（時間区間）の特徴ベクトル X_i とする。

【0056】

$$X_i = [P_m, P_{sd}, W_m, W_{sd}, \dots, G_m, G_{sd}] \quad (24)$$

図3には、上述したような18次元の特徴ベクトルを求める特徴抽出部4の具体的構成を示す。

【0057】

この図3において、端子10には、図1の時区間分割部3での時区間分割により得られた第 i ブロックのオーディオ信号 $s_i(t)$ が入力される。この第 i ブロック毎のオーディオ信号 $s_i(t)$ は、波形分析部11、スペクトル分析部12、LPC分析部13に送られる。

【0058】

波形分析部11では、当該第 i ブロックのオーディオ信号 $s_i(t)$ に対して、前記式(1)及び(2)にて説明した信号パワーの平均 P_m 及び標準偏差 P_{sd} を求める。上記信号パワーの平均 P_m と標準偏差 P_{sd} は、前記18次元の特徴ベクトル X_i のうちの2つとして、それぞれ対応する端子22、23から後段のベクトル量子化部5へ送られる。

【0059】

スペクトル分析部12では、上記第 i ブロックのオーディオ信号 $s_i(t)$ のスペ

クトル分析を行い、当該第 i ブロックの信号のスペクトログラム $S_i(t, w)$ を生成する。この第 i ブロックの信号のスペクトログラム $S_i(t, w)$ は、スレッシュホールド処理部 1 4、低周波成分抽出部 1 5、中間周波成分抽出部 1 6、高周波成分抽出部 1 7、ピッチ抽出部 1 8、調波構造度抽出部 1 9 に送られる。

【 0 0 6 0 】

上記スレッシュホールド処理部 1 4 では、当該第 i ブロックの信号のスペクトログラム $S_i(t, w)$ を用い、前記式 (3) 及び (4) にて説明したスペクトログラムの広がり W_m の平均 W_m 及び標準偏差 W_{sd} を求める。上記スペクトログラムの広がり W_m の平均 W_m 及び標準偏差 W_{sd} は、前記 1 8 次元の特徴ベクトル X_i のうちの 2 つとして、それぞれ対応する端子 2 4、2 5 から後段のベクトル量子化部 5 へ送られる。

【 0 0 6 1 】

上記低周波成分抽出部 1 5 では、当該第 i ブロックの信号のスペクトログラム $S_i(t, w)$ を用い、前記式 (6) 及び (7) にて説明した低周波成分のパワーの平均 L_m 及び標準偏差 L_{sd} を求める。上記低周波成分のパワーの平均 L_m 及び標準偏差 L_{sd} は、前記 1 8 次元の特徴ベクトル X_i のうちの 2 つとして、それぞれ対応する端子 2 6、2 7 から後段のベクトル量子化部 5 へ送られる。

【 0 0 6 2 】

上記中間周波成分抽出部 1 6 では、当該第 i ブロックの信号のスペクトログラム $S_i(t, w)$ を用い、前記式 (9) 及び (10) にて説明した中間周波成分のパワーの平均 M_m 及び標準偏差 M_{sd} を求める。上記中間周波成分のパワーの平均 M_m 及び標準偏差 M_{sd} は、前記 1 8 次元の特徴ベクトル X_i のうちの 2 つとして、それぞれ対応する端子 2 8、2 9 から後段のベクトル量子化部 5 へ送られる。

【 0 0 6 3 】

上記高周波成分抽出部 1 7 では、当該第 i ブロックの信号のスペクトログラム $S_i(t, w)$ を用い、前記式 (12) 及び (13) にて説明した高周波成分のパワーの平均 H_m 及び標準偏差 H_{sd} を求める。上記高周波成分のパワーの平均 H_m 及び標準偏差 H_{sd} は、前記 1 8 次元の特徴ベクトル X_i のうちの 2 つとして、それぞれ対応する端子 3 0、3 1 から後段のベクトル量子化部 5 へ送られる。

【 0 0 6 4 】

上記ピッチ抽出部 1 8 では、当該第 i ブロックの信号のスペクトログラム $S_i(t, w)$ を用い、前記式 (1 5) 及び (1 6) にて説明したピッチ周波数の平均 F_m 及び標準偏差 F_{sd} を求める。上記ピッチ周波数の平均 F_m 及び標準偏差 F_{sd} は、前記 1 8 次元の特徴ベクトル X_i のうちの 2 つとして、それぞれ対応する端子 3 2、3 3 から後段のベクトル量子化部 5 へ送られる。

【 0 0 6 5 】

上記調波構造度抽出部 1 9 では、当該第 i ブロックの信号のスペクトログラム $S_i(t, w)$ を用い、前記式 (1 8) 及び (1 9) にて説明した調波構造度の平均 A_m 及び標準偏差 A_{sd} を求める。上記調波構造度の平均 A_m 及び標準偏差 A_{sd} は、前記 1 8 次元の特徴ベクトル X_i のうちの 2 つとして、それぞれ対応する端子 3 4、3 5 から後段のベクトル量子化部 5 へ送られる。

【 0 0 6 6 】

また、上記 L P C 分析部 1 3 では、上記第 i ブロックのオーディオ信号 $s_i(t)$ の L P C 分析を行い、当該第 i ブロックにおける L P C 分析の残差信号 $r(t)$ を生成する。この第 i ブロックにおける L P C 分析の残差信号 $r(t)$ は、残差エネルギー抽出部 2 0、ピッチゲイン抽出部 2 1 に送られる。

【 0 0 6 7 】

上記残差エネルギー抽出部 2 0 では、当該第 i ブロックにおける L P C 分析の残差信号 $r(t)$ を用い、前記式 (2 0) 及び (2 1) にて説明した L P C 分析の残差エネルギーの平均 R_m 及び標準偏差 R_{sd} を求める。上記残差エネルギーの平均 R_m 及び標準偏差 R_{sd} は、前記 1 8 次元の特徴ベクトル X_i のうちの 2 つとして、それぞれ対応する端子 3 6、3 7 から後段のベクトル量子化部 5 へ送られる。

【 0 0 6 8 】

上記ピッチゲイン抽出部 2 1 では、当該第 i ブロックにおける L P C 分析の残差信号 $r(t)$ を用い、前記式 (2 2) 及び (2 3) にて説明した L P C 分析の残差信号のピッチゲインの平均 G_m 及び標準偏差 G_{sd} を求める。上記残差信号のピッチゲインの平均 G_m 及び標準偏差 G_{sd} は、前記 1 8 次元の特徴ベクトル X_i のうちの 2 つとして、それぞれ対応する端子 3 8、3 9 から後段のベクトル量子化部

5へ送られる。

【0069】

次に、上述したような18次元の特徴ベクトルが供給されたベクトル量子化部5では、ベクトル量子化法を用い、上記18次元の特徴ベクトルから各ブロック毎のオーディオ信号を分類することになるが、ここで、上記ブロック毎のオーディオ信号を分類する際の分類クラスについて以下に詳細に説明する。

【0070】

本実施の形態において、各ブロック毎のオーディオ信号を分類する際には、以下に述べるように、各ブロック毎に構造的な分類クラスと音源的な分類クラスによる分類を行うようにしている。

【0071】

先ず、上記構造的な分類クラスについて詳細に説明する。

【0072】

上記構造的な分類クラスとは、音源の種類によらず、ブロック中の信号の構造パターンに基づく分類を意味する。すなわち、本実施の形態では、上記構造的な分類クラスとして、以下に述べる無音構造(silent)、単音源構造(single)、二音源構造(double)、音源変化構造(change)、音源同時変化構造(multiple change)、音源部分変化構造(partial change)、その他の構造(other)の各構造的な分類パターンを定義する。なお、図4には、各構造的な分類パターンを模式的に示している。

【0073】

上記無音構造パターンとは、そのブロック中に有意な音が存在せず、無音100となっているパターンを意味する。

【0074】

上記単音構造パターンとは、そのブロック中の略々全域に渡り、有意な音101が1種類だけ存在するパターンを意味する。

【0075】

上記二音源構造パターンとは、そのブロック中の略々全域に渡り、有意な音が2種類（音102及び103）存在するパターンを意味し、例えばBGM（バック

クランドミュージック)の上に音声があるパターンなどである。

【0076】

上記音源変化構造パターンとは、そのブロック中で音源の種類が切り替わるパターンであり、例えば音声104から音楽105への切り替わりなどである。但し、有意な音から無音への変化、無音から有意な音への変化も当該音源変化構造パターンに含むことにする。

【0077】

上記音源同時変化構造パターンとは、そのブロック中で二音源音が同時に他の二音源音に切り替わる（例えば二音源音106及び108が他の二音源音107及び109に切り替わる）パターンを意味し、また、単音源音（又は無音）から二音源音へ切り替わる（例えば単音源音113が二音源音114及び115に切り替わる）パターンや、二音源音から単音源音（又は無音）へ切り替わる（例えば二音源110及び111が単音源112に切り替わる）パターンも、当該音源同時変化構造パターンに含まれる。当該音源同時変化構造パターンとしては、例えばBGMと音声が略々同時に終了する場合などを挙げることができる。

【0078】

上記音源部分変化構造パターンとは、そのブロック中、1種類の音（音118）は略々全域に渡り存在し、同時に存在する音が切り替わる（例えば音116が音117に切り替わる）パターンを意味し、二音源音が単音源音へと切り替わる場合も含む。この音源部分変化構造パターンとしては、例えばBGM継続中に、同時に存在していた音声のみが終了する場合などを挙げることができる。

【0079】

上記その他の構造パターンとは、上述の各構造には属さないパターンを意味し、例えば、そのブロック中に3つの音（例えば音119, 120, 121）が同時に存在するパターンや、そのブロック中に2回以上の音の切り替わり（例えば音122が音123に切り替わり、さらに音124に切り替わる）が存在するパターンなどを挙げることができる。

【0080】

次に、上記音源的な分類クラスについて詳細に説明する。

【0081】

上記音源的な分類クラスとは、音源の種類による分類を意味する。すなわち、本実施の形態では、以下に述べる音声、音楽、雑音、打撃音、環境音等の音源の種類によるクラス分類を行うようにしている。

【0082】

上記音声は、人間の音声の音であり、必要に応じて男性、女性、その他の音声（子供や加工音声など）に詳細分類することもできる。

【0083】

上記音楽は、いわゆる音楽の音であり、必要に応じて楽器音楽、ボーカル音楽、その他の音楽（ラップ音楽など）に詳細分類することもできる。

【0084】

上記雑音は、機械等から発生する白色性雑音である。

【0085】

上記打撃音は、例えばドアノック音、足音、拍手（少人数での拍手）等、主として打撃を発生源とし、その打撃の発生直後に音量が大きくなり、以後、減衰性を持つ音のクラスである。当該打撃音は、必要に応じて、発生源別に詳細分類することもできる。

【0086】

上記環境音は、雑踏音、拍手音（多人数での拍手）、歓声、エンジン音等、その他の音を全て含むクラスである。当該環境音は、必要に応じて、発生源別に詳細分類することもできる。

【0087】

図1のベクトル量子化部5では、前記18次元の特徴ベクトルを用いて、上述したような構造的な分類クラス及び音源的な分類クラスによる分類を行うことになる。

【0088】

以下、ベクトル量子化部5における上記特徴ベクトルを用いたブロック毎の分類法について詳細に説明する。

【0089】

本実施の形態では、ブロック毎のオーディオ信号の分類を行う場合に、図5のフローチャートに示すような三段階に分けた分類を行うようにしている。

【0090】

この図5において、ステップS1の処理として、上記第iブロックにて求めた18次元の特徴ベクトル X_i が入力されると、ベクトル量子化部5では、まず、ステップS2の処理として、無音分類 (Silent Classification)を行う。すなわち、ベクトル量子化部5は、例えば前記信号パワーの平均 P_m 及び標準偏差 P_{sd} を用い、それら信号パワーの平均 P_m 及び標準偏差 P_{sd} がそれぞれ一定閾値以下であるか否かを調べることにより、当該第iブロックのオーディオ信号が前記構造的な分類クラスのうち無音構造のパターンに分類されるか否かの判定を行う。

【0091】

当該ステップS2の判定処理において、無音構造であると判定した場合、ベクトル量子化部5は、ステップS6の処理として、無音構造パターンの分類結果を出力した後、ステップS1の処理に戻り、次のブロックについての処理に進む。一方、ステップS2の判定処理において、無音構造でないと判定した場合、ベクトル量子化部5の処理は、ステップS3の処理に移行する。

【0092】

ステップS3の処理に進むと、ベクトル量子化部5は、変化音分類 (Change Classification)を行う。ここでの分類では、ブロック内の信号が変化構造を有するか否か、すなわちブロック内の信号が音源変化構造 (change)、音源同時変化構造 (multiple change)、音源部分変化構造 (partial change) の何れのパターンであるか、或いはそれら以外の単音源音 (single)、二音源音 (double)、その他 (other) の無変化音のパターンであるかの二種に分類する。

【0093】

このような分類を行うため、ベクトル量子化部5では、まず、分類の対象とする第iブロックの前後の第i-1ブロックの特徴ベクトル X_{i-1} と第i+1ブロックの特徴ベクトル X_{i+1} を用いて、新たな特徴ベクトル Y_i を生成する。すなわち、下記式(25)により、上記新たな特徴ベクトル Y_i を求める。

【0094】

$$Y_i = (X_{i+1} - X_{i-1}) / (X_{i+1} + X_{i-1}) \quad (25)$$

なお、この式(25)での加減算及び除算は、特徴ベクトル X_{i-1} と特徴ベクトル X_{i+1} をそれぞれ構成する各特徴量毎に行う。

【0095】

ベクトル量子化部5は、上述のようにして新たな特徴ベクトル Y_i を求めると、当該新たな特徴ベクトル Y_i と、予め学習してあるVQ符号帳8との比較を行い、最もマハラノビス距離の近いセントロイドを探索し、当該最も近いセントロイドが代表しているカテゴリ（この場合は変化構造であるか否か）を求める。当該ステップS3の判定処理において、変化構造であると判定した場合、ベクトル量子化部5は、ステップS7の処理として、当該第iブロック内のオーディオ信号が音源変化構造、音源同時変化構造、音源部分変化構造の何れかのパターンであることを示す分類結果と、上記ベクトル量子化により得られたセントロイドに対する最小距離の逆数（構造的な分類の信頼度）を出力する。その後は、ステップS1の処理に戻り、次のブロックについての処理に進む。一方、ステップS3の判定処理において、変化構造でない（無変化音である）と判定した場合、ベクトル量子化部5の処理は、ステップS4の処理に移行する。

【0096】

ステップS4の処理に進むと、ベクトル量子化部5は、無変化音のパターンすなわち単音源音、二音源音、その他のパターンについての音源分類（source Classification）を行い、次のステップS5にて、音声、音楽、雑音、打撃音、環境音等の音源の種類によるクラス分類の分類結果を出力する。すなわち、ベクトル量子化部5は、ベクトル量子化法を用い、第iブロックについての18次の特徴ベクトル X_i と、予め学習してあるVQ符号帳8との比較を行い、最もマハラノビス距離の近いセントロイドを探索し、当該最も近いセントロイドが代表している分類ラベルと、このときのセントロイドに対する最小距離の逆数（そのカテゴリに対する分類の信頼度）を、分類結果として出力する。当該ステップS5の処理後、ベクトル量子化部5は、ステップS1の処理に戻り、次のブロックについての処理に進む。

【0097】

図6には、図1のベクトル量子化部5及びVQ符号帳8において、上記図5に示したフローチャートのステップS3及びステップS7の処理を実現する場合の機能的な構成を表し、図7には、図1のベクトル量子化部5及びVQ符号帳8において、上記図5のフローチャートのステップS4及びステップS5の処理を実現する場合の機能的な構成を表す。すなわち、上記図5のステップS3及びステップS7の処理を実現する場合、図1のベクトル量子化部5及びVQ符号帳8の機能は、図6のように表すことができ、また、上記図5のステップS4及びステップS5の処理を実現する場合、図1のベクトル量子化部5及びVQ符号帳8の機能は、図7のように表すことができる。なお、図6と図7は、ベクトル量子化部5内の処理を分かり易く説明するために、2つの図として分けて描いているが、実際のベクトル量子化部5の構成は一つであり、図5のフローチャートのステップS2やステップS3の判定結果に応じて、その内部の機能が図6や図7に示すように動作するものである。

【0098】

先ず図6において、ベクトル量子化部5の端子51には、分類の対象とする第 i ブロックの前のブロックである第 $i-1$ ブロックの特徴ベクトル X_{i-1} が供給され、端子52には第 i ブロックの後のブロックである第 $i+1$ ブロックの特徴ベクトル X_{i+1} が供給される。これら第 $i-1$ ブロックの特徴ベクトル X_{i-1} と第 $i+1$ ブロックの特徴ベクトル X_{i+1} は、ベクトル量子化部5内の特徴混合演算部53に送られる。

【0099】

当該特徴混合演算部53は、前記式(25)に示した特徴混合演算式を用いて、それら第 $i-1$ ブロックの特徴ベクトル X_{i-1} と第 $i+1$ ブロックの特徴ベクトル X_{i+1} を混合し、新たな特徴ベクトル Y_i を生成する。当該新たな特徴ベクトル Y_i は、ベクトル量子化部5の主要部としての距離計算比較演算部54に送られる。

【0100】

距離計算比較演算部54では、当該新たな特徴ベクトル Y_i と前記VQ符号帳

8との比較を行い、上記新たな特徴ベクトル Y_i と最もマハラノビス距離に近いセントロイドを探索し、そのセントロイドが代表しているカテゴリを分類結果（変化音か又は無変化音か）として出力する。この分類結果を表す記述子は、当該ベクトル量子化部5の出力端子55から出力されることになる。

【0101】

一方、図7において、ベクトル量子化部5の端子61には、分類の対象である第 i ブロックの特徴ベクトル X_i が供給される。当該第 i ブロックの特徴ベクトル X_i は、ベクトル量子化部5の主要部としての距離計算比較演算部62に送られる。

【0102】

距離計算比較演算部62では、当該特徴ベクトル X_i と前記VQ符号帳8との比較を行い、上記特徴ベクトル X_i と最もマハラノビス距離に近いセントロイドを探索し、そのセントロイドが代表しているカテゴリを分類結果（ここでは、音声、音楽、雑音、環境音等）として出力する。この分類結果を表す記述子は、当該ベクトル量子化部5の出力端子63から出力されることになる。

【0103】

以上説明したように、本発明の第1の実施の形態の信号処理装置によれば、長時間に渡り様々な音源が様々なパターンを示す時系列信号であるオーディオ信号、すなわち、音声や音楽、環境音、雑音などの音が同時的若しくは継時的に重畳された音の信号に関して、ブロック毎に信号の音源の種類及び構造の種類を分類することが可能となっている。また、本実施の形態の信号処理装置によれば、音声区間などを特定することができるため、例えば音声認識や音響信号符号化等の前処理として用いることで、任意のオーディオ信号に対して、適切な認識方法や符号化方法を自動的に選択することが可能となる。

【0104】

次に、本発明の第2の実施の形態について説明する。

【0105】

ここで、蓄積された長いオーディオ信号のストリームから、例えば必要な部分を検索するような場合、一般的には、例えば早送り再生を行いながら試聴し、必

要な場所から通常再生を開始するといったことが行われている。しかしながら、このような検索手法を用いたのでは、所望のオーディオ信号を見つけるまでに非常に長時間を要するのみならず、ユーザは早送り再生されている音を聞き続けるといった多大な負担を強いられることになる。

【0106】

本発明の第2の実施の形態では、前述した第1の実施の形態にて説明したオーディオ信号の分類手法による分類結果のうち、前記変化構造の分類結果（特に音源変化構造、音源同時変化構造）を用いて、オーディオ信号の切り替わり点（以下、シーンチェンジと呼ぶ）を検出し、当該シーンチェンジを検出した後に無音構造からそれ以外の構造に変化した時点で通常再生を開始することにより、所望のオーディオ信号の検索を容易にする例について説明する。

【0107】

図8には、第1の実施の形態のオーディオ信号分類手法による分類結果のうち、前記変化構造の分類結果（特に音源変化構造、音源同時変化構造）を用いて、オーディオ信号のシーンチェンジを検出し、所望のオーディオ信号を容易に検索可能とした、本発明の第2の実施の形態の信号処理装置の概略構成を示す。また、図9には、当該図8の構成における処理の流れを示す。

【0108】

以下、これら図8及び図9を用いて第2の実施の形態の信号処理装置の構成及び動作の説明を行う。

【0109】

この図8において、再生部71は、様々な情報記録媒体や通信媒体を介してオーディオ信号を再生し、出力可能であると共に、後述する再生コントロールブロック77によりその再生動作が制御されるものである。ここで、本実施の形態において所望のオーディオ信号の検索を行う場合には、当該再生部71から例えば早送り再生によるオーディオ信号が出力され、その早送り再生によるオーディオ信号は前記第1の実施の形態の信号処理装置に対応する分類部74に入力される。

【0 1 1 0】

分類部 7 4 では、ステップ S 1 1 の処理として、上記早送り再生によるオーディオ信号に対して、前述したようなブロック化、特徴抽出、ベクトル量子化を用いた分類処理を行い、当該ブロックの分類結果を表す記述子（タグ）を出力する。当該記述子は、後段のシーンチェンジ検出部 7 5 に送られる。

【0 1 1 1】

なお、予めオーディオ信号に対して当該分類部 7 4 における分類処理を行い、オーディオ信号に同期させて上記記述子を付加しておくようにし、当該記述子が付加されたオーディオ信号を上記再生部 7 1 から出力するようなことも可能である。但し、このように予め記述子を付加したオーディオ信号を再生部 7 1 から出力するようにした場合は、上記分類部 7 4 での分類処理はスキップされ、上記記述子が上記シーンチェンジ検出部 7 5 に直接入力されることになる。

【0 1 1 2】

シーンチェンジ検出部 7 5 では、ステップ S 1 2 の処理として、上記分類部 7 4 から上記ブロックの分類結果を表す記述子を受け取ると、ステップ S 1 3 の処理として、その記述子に基づいて、当該ブロックのオーディオ信号が前述した音源変化構造（change）、又は音源同時変化構造（multiple change）であるか否かチェックする。

【0 1 1 3】

ここで、当該ステップ S 1 3 において、当該ブロックの信号が音源変化構造又は音源同時変化構造の何れにも該当しないと判定された場合、シーンチェンジ検出部 7 5 は、再生コントロール部 7 7 に対してその旨の検出結果信号を出力する。この時の検出結果信号を受け取ると、再生コントロール部 7 7 は、上記再生部 7 1 を制御し、上記早送り再生を続けさせる。これにより、当該図 8 の装置における処理は、ステップ S 1 1 に戻り、次のブロックのオーディオ信号に対してステップ S 1 1 からステップ S 1 3 までの処理が行われる。

【0 1 1 4】

一方、ステップ S 1 3 において、その時点でのブロックの信号が音源変化構造又は音源同時変化構造の何れかに該当すると判定された場合、シーンチェンジ検

出部 7 5 は、再生コントロール部 7 7 に対してその旨の検出結果信号を出力する。この時の検出結果信号を受け取ると、再生コントロール部 7 7 は、上記再生部 7 1 を制御し、上記早送り再生を続けさせる。これにより、分類部 7 4 では、ステップ S 1 4 の処理として、上記早送り再生による次のブロックのオーディオ信号に対して、前述したようなブロック化、特徴抽出、ベクトル量子化を用いた分類処理を行い、当該次のブロックの分類結果を表す記述子を出力する。

【0 1 1 5】

シーンチェンジ検出部 7 5 では、ステップ S 1 5 の処理として、上記ステップ S 1 4 による次のブロックについての記述子を受け取ると、ステップ S 1 6 の処理として、その記述子に基づいて、当該次のブロックのオーディオ信号が前述した無音構造 (silent) であるか否かチェックする。

【0 1 1 6】

ここで、当該ステップ S 1 6 において、当該次のブロックの信号が無音構造であると判定された場合、シーンチェンジ検出部 7 5 は、再生コントロール部 7 7 に対してその旨の検出結果信号を出力する。この時の検出結果信号を受け取ると、再生コントロール部 7 7 は、上記再生部 7 1 を制御し、上記早送り再生を続けさせる。これにより、当該図 8 の装置における処理は、ステップ S 1 4 に戻り、さらに次のブロックのオーディオ信号に対してステップ S 1 4 からステップ S 1 6 までの処理が行われる。

【0 1 1 7】

一方、ステップ S 1 6 において、その時点でのブロックの信号が無音構造に該当しないと判定された場合、シーンチェンジ検出部 7 5 は、再生コントロール部 7 7 に対してその旨の検出結果信号を出力する。この時の検出結果信号を受け取ると、再生コントロール部 7 7 は、上記再生部 7 1 を制御し、上記早送り再生を停止し、通常速度での再生を行わせる。当該通常速度にて再生されたオーディオ信号は、混合部 7 2 を介し、さらに端子 7 3 を介して図示しないディスプレイ装置等のスピーカに送られる。これにより、ディスプレイ装置等のスピーカからは、通常速度で再生されたオーディオ信号に基づく音が放音されることになる。

【0 1 1 8】

すなわち、ステップ S 1 6 においてその時点でのブロックの信号が無音構造に該当しないと判定された場合、当該ブロックのオーディオ信号は新たなシーンの音の信号であると考えられるため、本実施の形態装置では当該新たなシーンの開始部分のオーディオ信号を通常速度で再生する。これにより、ユーザは当該通常速度で再生された音によりシーンチェンジ後の音が所望の音であるか否か判断できることになる。また、ステップ S 1 6 において、ブロックの信号が無音構造に該当しないと判定された場合、シーンチェンジ検出部 7 5 からの検出結果信号は報知信号生成部 7 6 にも送られる。当該報知信号生成部 7 6 は、上記検出結果信号を受けると、シーンチェンジを検出したことをユーザに知らせるための報知音信号を生成して出力する。当該報知音信号は、混合器 7 2 を介してディスプレイ装置等のスピーカに送られ、したがってスピーカからは上記シーンチェンジの検出に対応する報知音が放音され、これにより、ユーザはシーンチェンジが検出されたことを認識可能となる。なお、報知信号生成部 7 6 が出力する報知信号は、ディスプレイ装置の画面上にシーンチェンジが検出された旨を表す表示信号とすることもできる。但し、当該表示信号としての報知信号を報知信号生成部 7 6 から出力するようにした場合、その報知信号は混合器 7 2 ではなくディスプレイ装置の表示部に送られることになる。

【0.1 1 9】

以上説明したように、本発明の第 2 の実施の形態の信号処理装置によれば、第 1 の実施の形態の場合と同様にしてオーディオ信号を分類した分類結果に応じて、オーディオ信号の変化点（シーンチェンジ）を検索できるため、例えば話題の切り替わる部分や、放送番組の切り替わる時刻等の自動検出が可能となり、これにより、マルチメディアデータの検索が容易となる。また、当該第 2 の実施の形態の信号処理装置によれば、早送り再生など人間に負担のかかる方法で全ての内容をチェックすることなく、シーンチェンジの開始部分のような検索に必要とされる可能性の高い部分の候補のみを通常速度で試聴し、選択することができることになる。

【0 1 2 0】

さらに、本発明の第2の実施の形態において、例えば映像のカットチェンジ（映像におけるカメラの切り替わり点）検出等と併用することで、いわゆるシーンチェンジ検出（映像上のまとまりを構成する単位に区切ること、一般に複数のカットからなる）の精度を向上させることも可能となる。

【0 1 2 1】

なお、本発明の第1、第2の実施の形態ではオーディオ信号を例に挙げたが、その他の信号、例えばビデオ信号などについても本発明は適用可能であり、それら他の信号についての分類、記述子の作成、検索等が可能となる。

【0 1 2 2】

【発明の効果】

以上の説明で明らかなように、本発明の信号分類方法及び装置においては、入力信号をブロック化してそのブロック内の信号から信号属性に関する特徴量を抽出し、その特徴量に基づいてブロック毎に信号のカテゴリ分類を行うことにより、入力信号を迅速且つ正確に分類可能であり、したがって、例えば長時間に渡り様々な音源が様々なパターンを示す時系列信号であるオーディオ信号を、ブロック毎に音源の種類及び構造の種類に分類することが可能である。

【0 1 2 3】

また、本発明の記述子作成方法及び装置においては、そのカテゴリ分類に応じた記述子を生成することにより、例えば、音声区間などを特定することが容易となり、例えば音声認識や音響信号符号化等の前処理として用いることで、任意のオーディオ信号に対して、適切な認識方法や符号化方法を自動的に選択することが可能となる。

【0 1 2 4】

また、本発明の信号検索方法及び装置においては、そのカテゴリ分類結果或いは記述子に基づいて信号検索を行うことにより、例えば音源の変化点が分類及び検索できるため、例えば話題の切り替わる部分や、放送番組の切り替わる時刻等の自動検出が可能となり、これによってマルチメディアデータの検索が容易となる。さらに、本発明の信号検索方法及び装置によれば、例えば映像をカットチェ

ンジ検出等と併用することで、いわゆるシーンチェンジ検出の精度を向上させることも可能である。

【図面の簡単な説明】

【図 1】

本発明の第 1 の実施の形態の信号処理装置の概略構成例を示すブロック回路図である。

【図 2】

オーディオ信号のブロック化の説明に用いる図である。

【図 3】

特徴抽出部の具体的構成例を示すブロック回路図である。

【図 4】

構造的な分類カテゴリの説明に用いる図である。

【図 5】

ベクトル量子化部で行われるブロック毎の分類処理の流れを示すフローチャートである。

【図 6】

ブロック内のオーディオ信号が変化音又は無変化音の何れであるか分類する際のベクトル量子化部の機能を構成として表す機能ブロック図である。

【図 7】

ブロック内のオーディオ信号を音声、音楽、雑音、環境音等に分類する際のベクトル量子化部の機能を構成として表す機能ブロック図である。

【図 8】

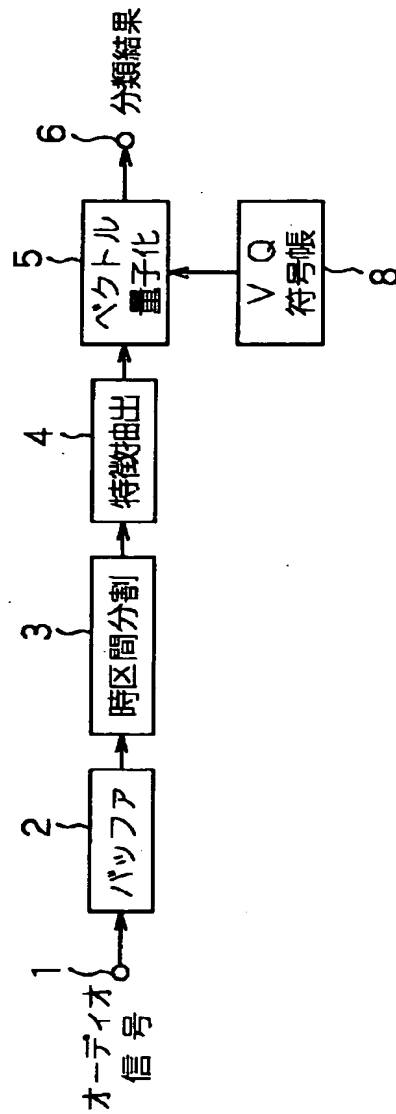
本発明の第 2 の実施の形態の信号処理装置の概略構成例を示すブロック回路図である。

【図 9】

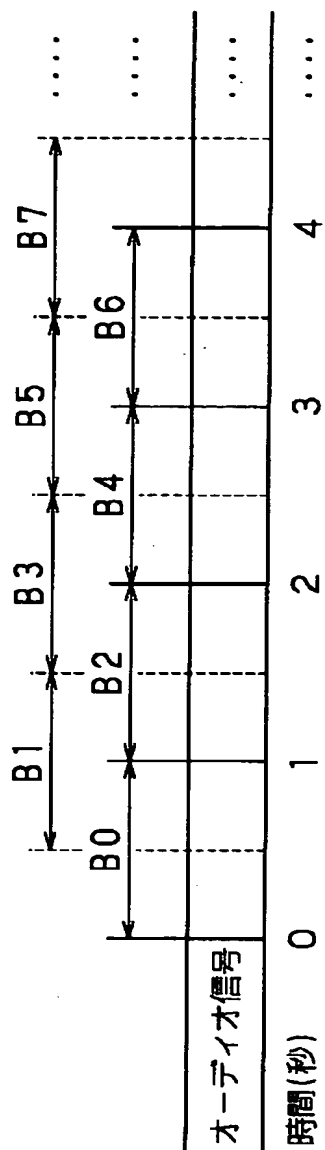
第 2 の実施の形態においてオーディオ信号のシーンチェンジを検出して所望のオーディオ信号を検索する際の処理の流れを示すフローチャートである。

【書類名】 図面

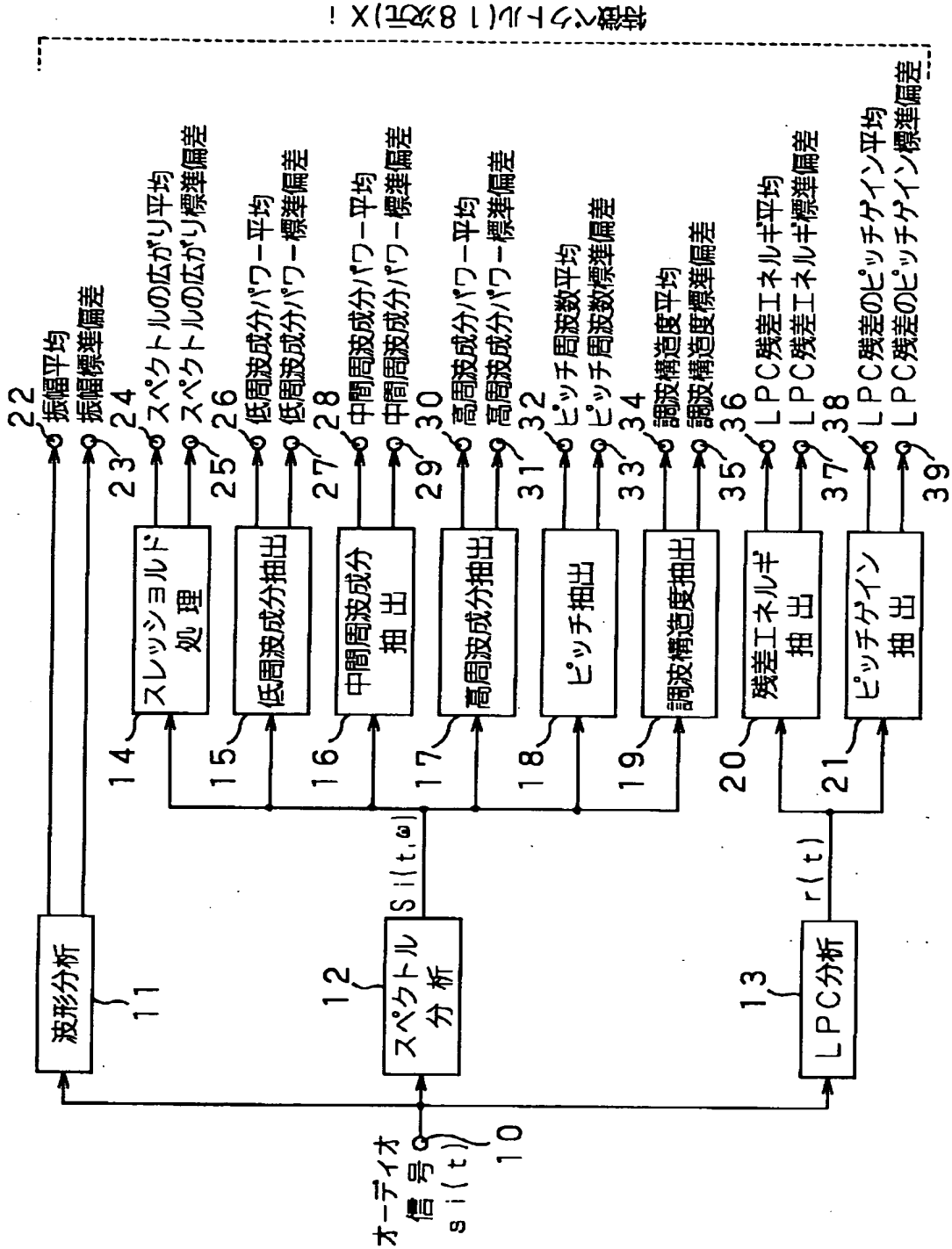
【図 1】



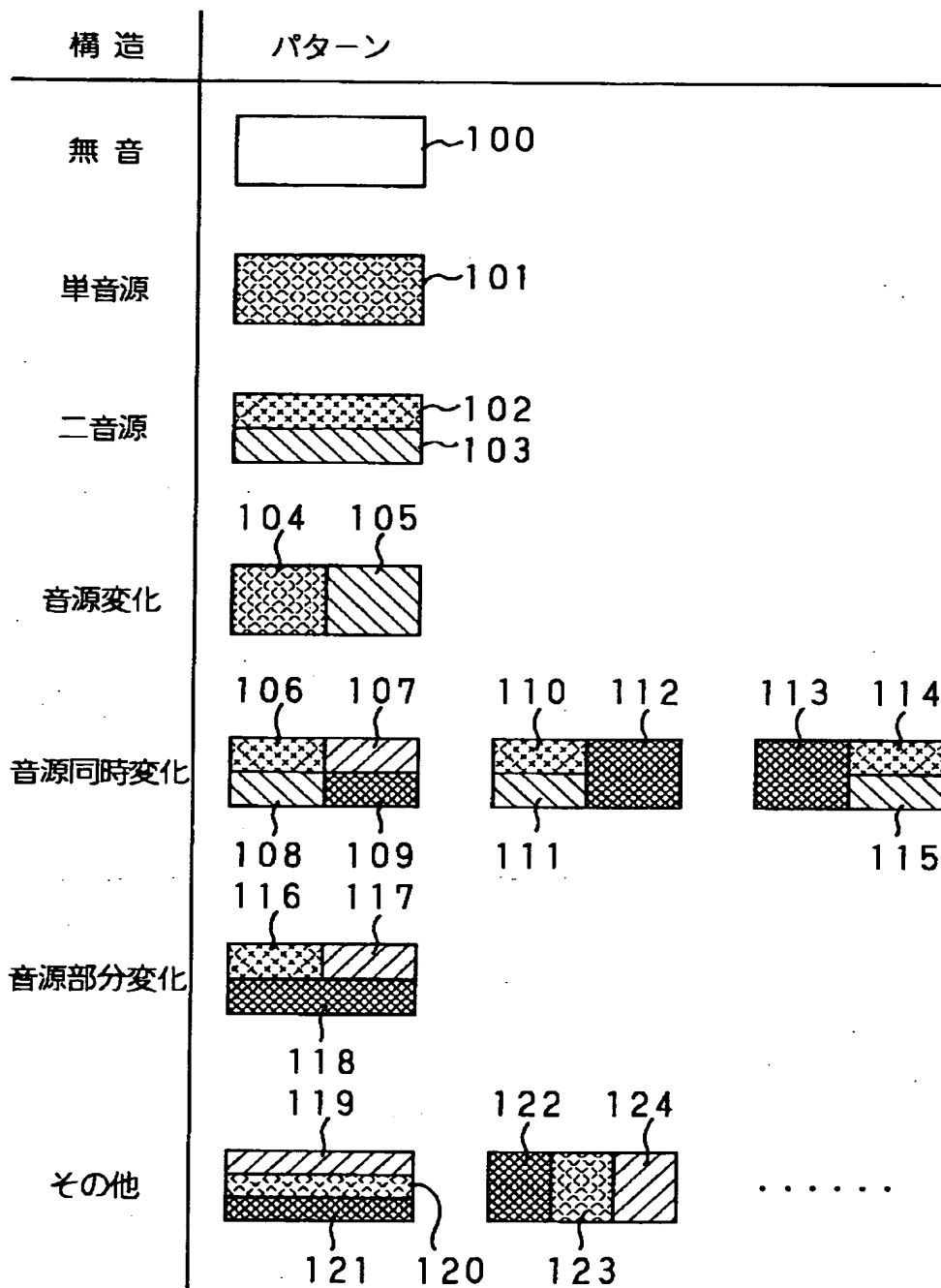
【図 2】



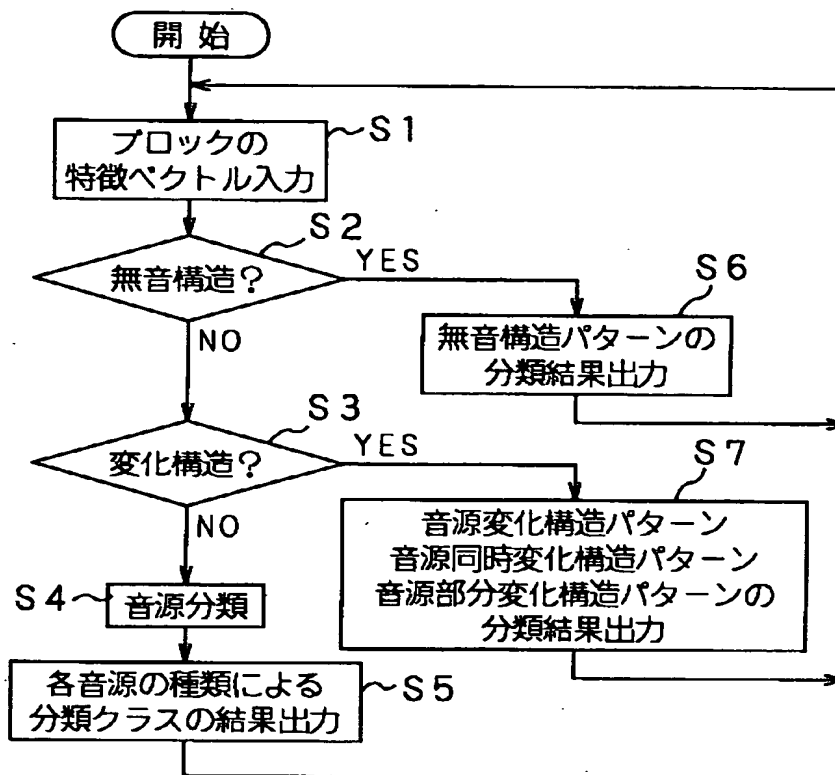
【図 3】



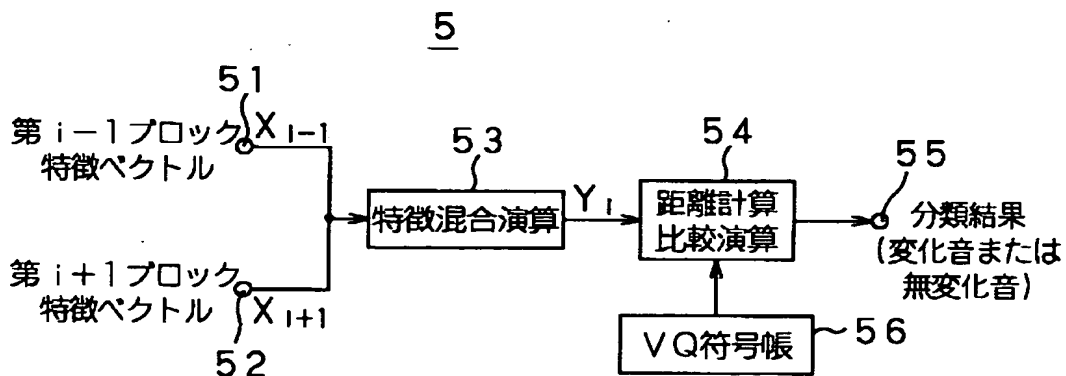
【図 4】



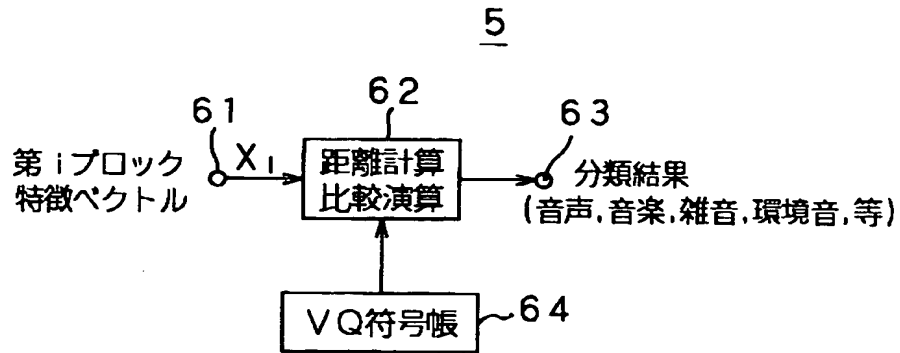
【図 5】



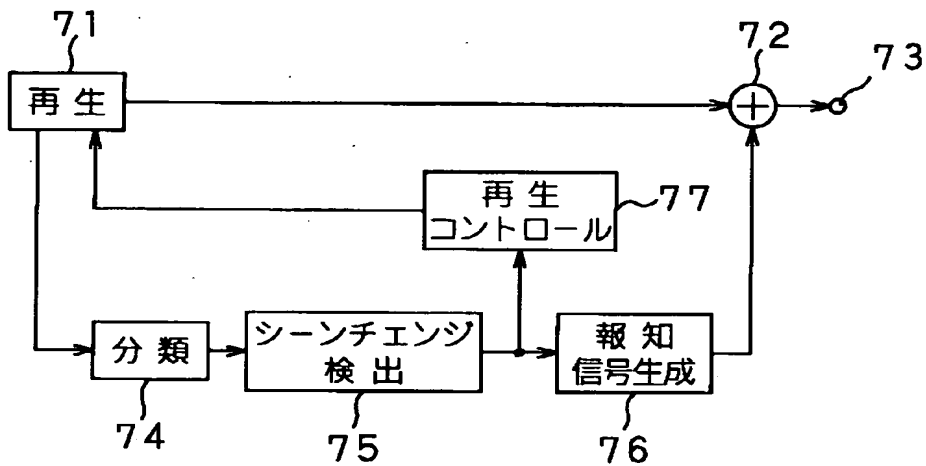
【図 6】



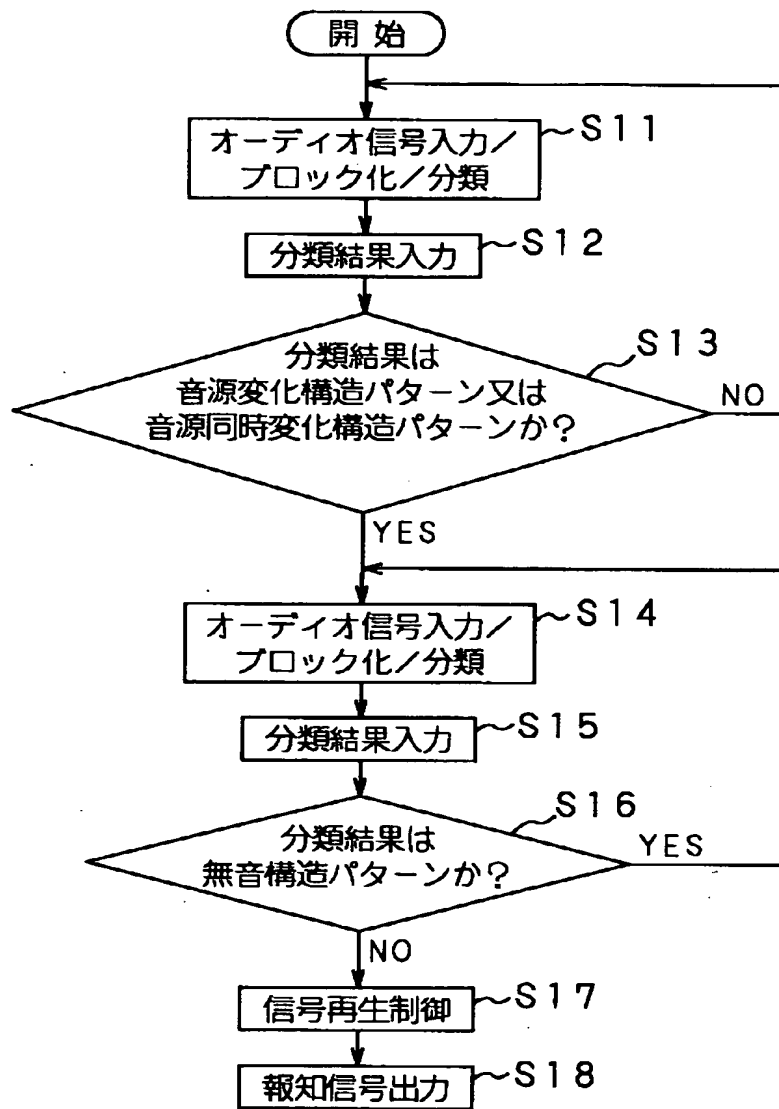
【図 7】



【図 8】



【図 9】



【書類名】 要約書

【要約】

【課題】 入力信号を迅速且つ正確に分類し、その分類に応じた記述子を生成可能とし、さらにその分類結果或いは記述子に基づいて入力信号の検索を可能とする。

【解決手段】 オーディオ信号を例えば 1 秒毎にブロック化する時区間分割ブロック 3 と、そのブロック内のオーディオより信号属性に関する 1 8 次元の特徴量を抽出する特徴抽出部 4 と、VQ 符号帳 8 と 1 8 次元の特徴量からなる特徴ベクトルを用いたベクトル量子化法により、ブロック毎にオーディオ信号のカテゴリ分類を行うベクトル量子化部 5 とを有する。ベクトル量子化部 5 からは、分類結果である分類ラベルとその信頼度が記述子として出力される。後段にて信号検索を行う場合には、分類結果或いは記述子を用いて信号検索を行う。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号 [000002185]

1. 変更年月日 1990年 8月30日

[変更理由] 新規登録

住 所 東京都品川区北品川6丁目7番35号

氏 名 ソニー株式会社